survxai: an R package for structure-agnostic explanations of survival models

Aleksandra Grudziaz[1, 2], **Alicja Gosiewska**[1], and **Przemyslaw Biecek**[1, 2]

1 Faculty of Mathematics and Information Science, Warsaw University of Technology **2** Faculty of Mathematics, Informatics, and Mechanics, University of Warsaw

Introduction

Predictive models are widely used in supervised machine learning. Three most common classes of such models are: *regression models*, where the target variable is continuous numeric, *classification models*, where the target variable is binary or categorical and *survival models*, where the target is some censored variable. Common examples of censored variables are times to death (but some cases lived for x years and are still alive), cessation of service by customer, or failure of machine components.

Modern survival models are often complex in structure; for example survival neural networks (Eleuteri, Tagliaferri, Milano, De Placido, & De Laurentiis, 2003) or survival random forest (Ishwaran, Kogalur, Blackstone, & Lauer, 2008). These models may be described by thousands of coefficients. Often such flexibility leads to high performance, but makes models opaque and hard to understand. This is acceptable in cases where only the model accuracy is important, but in cases that involve human decisions, it may not be informative enough. To trust model predictions one needs to see which features are important and how model predictions would change if some feature was changed.

The area of model interpretability or explanability has quickly gained the attention of machine learning experts. Understanding of complex models leads not only to higher trust in model predictions but also better models. Better, means that the models are more robust and obtains higher accuracy on validation data. See examples in the `DALEX` (Biecek, 2018) or `iml` (Molnar, 2018) R packages.

Existing tools for model agnostic explanations are focused on regression models and classification problems, as in both cases model predictions that may be summarised by a single number. Survival models require different approach as predictions are in a form of survival functions. Demand for such explainers has led to some model specific solutions, like `iSurvive` (Dempsey et al., 2017) for continuous time hidden Markov models. Yet, there is currently a lack of structure agnostic tools for survival models.

The `survxai` fills this gap. This R package is designed to deliver local and global explanations for survival models, in a structure-agnostic fashion. In the package documentation we demonstrate examples for survival random forest models and for Cox models. The `survxai` package consists of new implementations and visualisations of explainers, designed for survival models. Functions are well documented and the package is supplemented with unit tests, and illustrations. Regardless of the complexity of the model, the methods implemented in the `survxai` package maintain a certain level of interpretability, important in medical applications (Collett, 2015), churn analysis (Lu & Park, 2003) and others.

Explanations of survival models

The R package `survxai` is a tool for creating explanations of survival models. It is structure-agnostic, and thus works for both complex and simple survival models. It also allows for comparisons between two or more models.

Currently, four classes of model explainers are implemented. Two for local explanations (for a single prediction), and two for global explanations (for a whole model and population).

The package `survxai` is available on CRAN. It can be install using the command `install.packages('survxai')`. The development version of the package can be found at https://github.com/MI2DataLab/survxai.

Local methods are the explanations of a single observation.

- **The Ceteris Paribus** profile presents model responses around a single point in the feature space (Biecek, 2018). See Figure 1 for an example. Each panel is related to a single variable. Each single panel shows how a model prediction (survival curve) would change if only a single variable were changed. It is useful for *what-if* reasoning. Each curve in a panel is related to a different value of the selected variable. The Ceteris Paribus profile illustrates how the survival curve may change with the changing of values of variable.
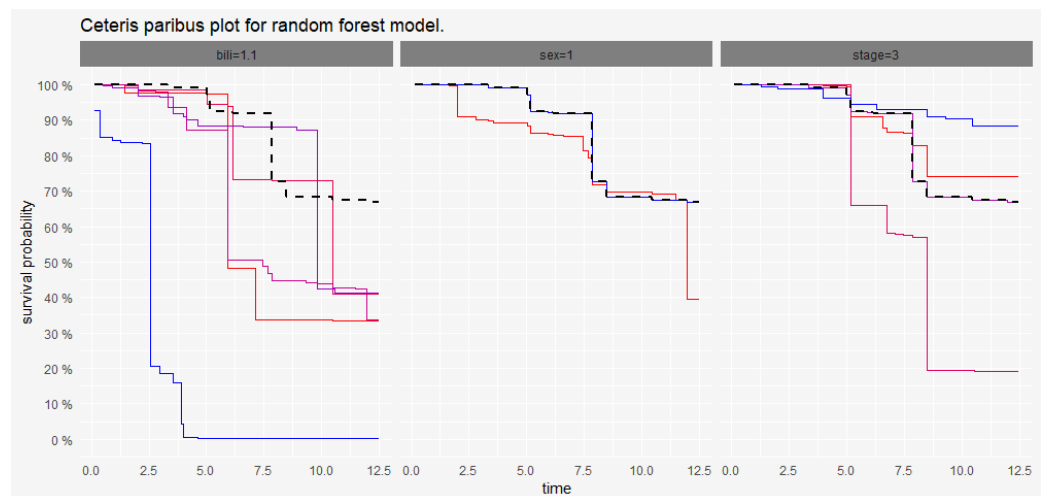

Figure 1: Ceteris Paribus plot for survival random forest model with three variables. The black dashed survival curve corresponds to an observation of interest. The left panel shows the survival curves for different values of bilirubin. Colors correspond to mean survival curves of observations from quintiles. From red which is the first quintile to blue which is the last one. The middle panel shows that prediction for sex=0 is worse than for sex=1 for times less than 7.5. The right panel analogously shows survival curves for different levels of the variable stage.

- **The Break Down** plot presents variable contributions to a model prediction (Staniak & Biecek, 2018). See Figure 2 for an example. The Break Down of predictions for survival models help to understand which factors drive survival probabilities for a single observation.
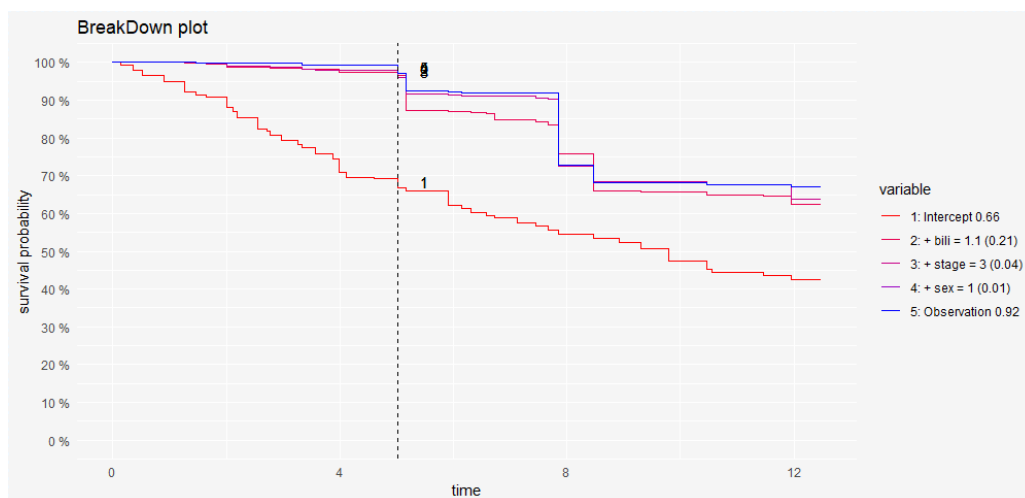
Figure 2: Break Down plot for survival random forest model. Variables bili and stage have the highest impact on the final prediction.

Global methods are explanations for performance and model structure.

- **The Variable Response** plot is designed to better understand the relation between a variable and a model output. See Figure 3 for an example. The variable response plot illustrates how the mean survival curve changes along with the changing values of the variable. It is inspired by Partial Dependence Plots (Greenwell, 2017).
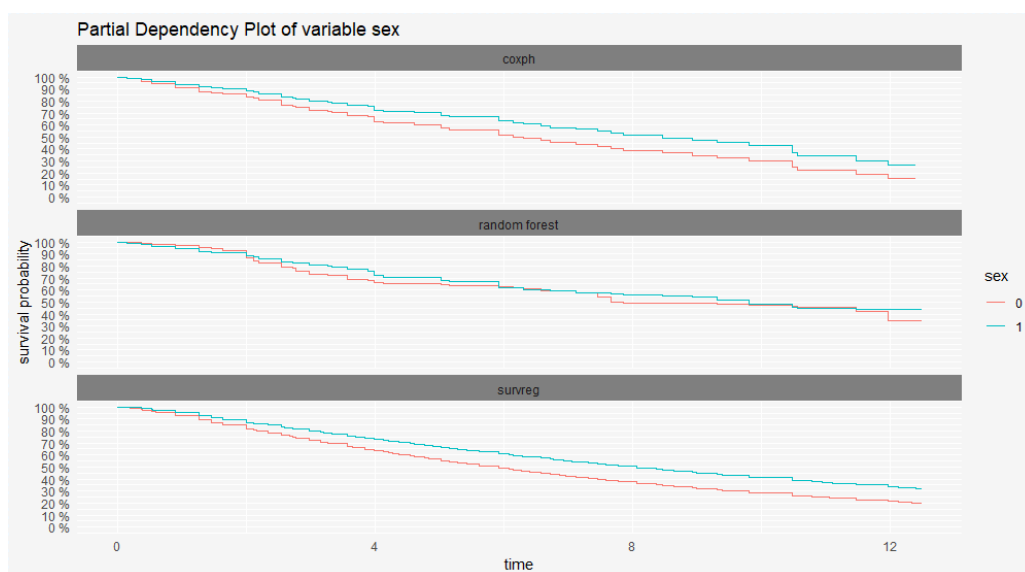

Figure 3: Variable response plots for three models and variable sex. In survival random forest, the sex variable affects model predictions in a different way than in other models.

- **The Model Performance** curves present prediction error for the chosen survival model, depending on time. See Figure 4 for an example. For computing prediction error, we use the expected Brier Score (Mogensen, Ishwaran, & Gerds, 2012). At a given time point t, the Brier score for a single observation is the squared difference between observed survival status and a model-based prediction of surviving time t.
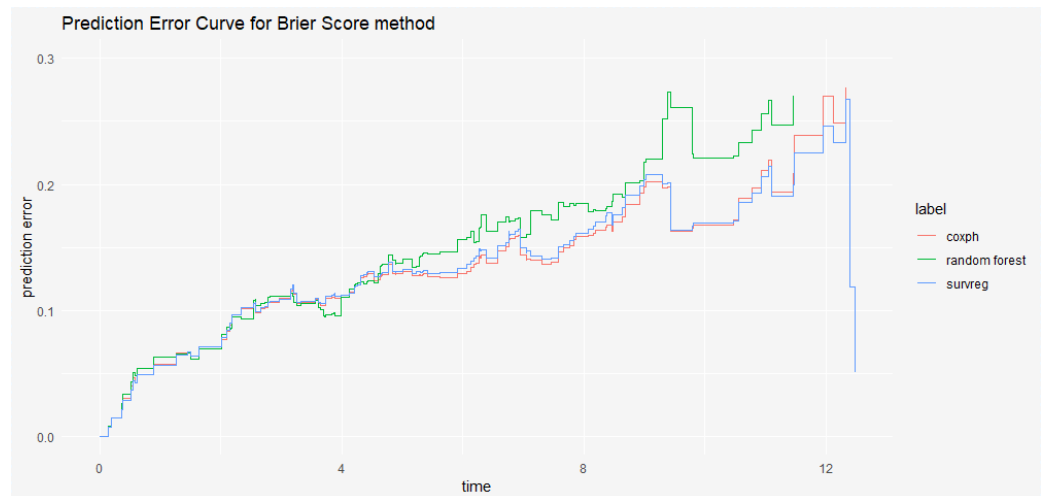
Figure 4: Model performance plots for three models. In random forest model, predictions are less accurate after year 4.

Conclusions and future work

Explainers implemented in the `survxai` package allow for exploration of one or more models in a feature-by-feature fashion. This approach will miss interactions between variables that may be handled by the models. The main problem with interactions is that number of interactions grows rapidly with the number of features what makes it hard to present in a readable form.

Acknowledgments

References

Biecek, P. (2018). DALEX: explainers for complex predictive models. Retrieved from http://arxiv.org/abs/1806.08915

Biecek, P. (2018). *CeterisParibus: Ceteris paribus plots (what-if plots) for a single observation.* Retrieved from http://CRAN.R-project.org/package=ceterisParibus

Collett, D. (2015). *Modelling survival data in medical research, third edition.* Chapman & hall/crc texts in statistical science. CRC Press. Retrieved from http://books.google.pl/books?id=Okf7CAAAQBAJ

Dempsey, W. H., Moreno, A., Scott, C. K., Dennis, M. L., Gustafson, D. H., Murphy, S. A., & Rehg, J. M. (2017). ISurvive: An interpretable, event-time prediction model for mHealth. In D. Precup & Y. W. Teh (Eds.), *Proceedings of the 34th international conference on machine learning*, Proceedings of machine learning research (Vol. 70, pp. 970–979). International Convention Centre, Sydney, Australia: PMLR. Retrieved from http://proceedings.mlr.press/v70/dempsey17a.html

Eleuteri, A., Tagliaferri, R., Milano, L., De Placido, S., & De Laurentiis, M. (2003). A novel neural network-based survival analysis model. *Neural Networks*, *16*(5), 855–864. doi:10.1016/S0893-6080(03)00098-4

Greenwell, B. M. (2017). pdp: An R Package for Constructing Partial Dependence Plots. *The R Journal*, *9*(1), 421–436. Retrieved from http://journal.r-project.org/archive/2017/RJ-2017-016/index.html

Ishwaran, H., Kogalur, U. B., Blackstone, E. H., & Lauer, M. S. (2008). Random survival forests. *The Annals of Applied Statistics*, *2*(3), 841–860. doi:doi:10.1214/08-AOAS169

Lu, J., & Park, O. (2003). Modeling customer lifetime value using survival analysis—an application in the telecommunications industry. *Data Mining Techniques*, 120–128.

Mogensen, U. B., Ishwaran, H., & Gerds, T. A. (2012). Evaluating random forests for survival analysis using prediction error curves. *Journal of Statistical Software*, *50*(11), 1–23. Retrieved from http://www.jstatsoft.org/v50/i11/

Molnar, C. (2018). Iml: An r package for interpretable machine learning. *Journal of Open Source Software*, *3*(26), 786. doi:10.21105/joss.00786

Staniak, M., & Biecek, P. (2018). Explanations of model predictions with live and break-Down packages. Retrieved from http://arxiv.org/abs/1804.01955